Reinforcement Learning: An Introduction

Solutions: Chapter 6

Mrinank Sharma

April 4, 2020

Exercise 6.1: TD to MC error

We have:

$$\delta_t = R_{t+1} + \gamma V_t(S_{t+1}) - V_t(S_t), \tag{1}$$

i.e., the error using the estimates at time t. We can write the update rule:

$$\begin{aligned} V_{t+1}(S_t) &= V_t(S_t) + \alpha[R_{t+1} + \gamma V_t(S_{t+1}) - V_t(S_t)] \\ &= V_t(S_t) + \alpha\delta_t. \end{aligned} \tag{2}$$

Rearranging yields:

$$V_{t+1}(s) - V_t(s) = \begin{cases} \alpha\delta_t, & s = S_t \\ 0, & \text{otherwise} \end{cases}. \tag{3}$$

Now, we follow the same steps as in the book.

$$\begin{aligned} G_t - V_t(S_t) &= R_{t+1} + \gamma G_{t+1} - V_t(S_t) \\ &= R_{t+1} + \gamma V_t(S_{t+1}) - V_t(S_t) + \gamma[G_{t+1} - V_{t+1}(S_{t+1}) + V_{t+1}(S_{t+1}) - V_t(S_{t+1})] \\ &= \delta_t + \gamma\alpha\mathbb{I}\{S_{t+1} = S_t\} + \gamma[G_{t+1} - V_{t+1}(S_{t+1})] \\ &\vdots \\ &= \sum_{k=t}^{T-1} \gamma^{k-t}\delta_k[1 + \alpha\gamma\mathbb{I}\{S_{k+1} = S_k\}]. \end{aligned} \tag{4}$$

This is incredibly similar as before, but we have additional $\alpha\gamma$ terms if the new state was the same as the old one, because only in these do we need our correction term.